

Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program

Pierre Borgnat, Eric Fleury, Céline Robardet, Antoine Scherrer

▶ **To cite this version:**

Pierre Borgnat, Eric Fleury, Céline Robardet, Antoine Scherrer. Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program. ECCS'09, Sep 2009, Warwick, United Kingdom. ensl-00408150

HAL Id: ensl-00408150

https://ens-lyon.hal.science/ensl-00408150

Submitted on 29 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noname manuscript No.
(will be inserted by the editor)

Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program

Pierre Borgnat · Eric Fleury · Céline Robardet · Antoine Scherrer

Received: date / Accepted: date

Abstract Public transportation systems like Lyon's bicycle community sharing program can be considered as a complex system composed of interconnected stations that exchange bicycles. Such system generates digital footprints that reveals the activity in the city over time and space and make possible their analyze. In this paper, the analysis deals with the spatial understanding and visualization of bicycle trips. We first study the activity in each station separately and then identify the main characteristics of the flow between stations.

Keywords Complex System · Community bicycle program · Vélo'v · Dynamic network · Spatial data.

1 Introduction

Community shared bicycle programs have been under great development in the recent past years all over Europe. We study Lyon's shared bicycle system, called *Vélo'v* and operated by the JCDecaux agency [1]. This program is a major one of its kind, having started in May 2005. Besides its evident interest as a new means to think about public transportation, such community shared programs offer a new way to look into the dynamics of travels inside a city, and more generally into its activity. The objective in this paper is to study the spatial patterns of activity from all the trips made with *Vélo'v*, going from an empirical point of view that can be compared to previous studies of equivalent systems in Paris (the *Vélib'* program studied in [2]) or in Barcelona (studied in [3]), to a more quantitave point of view of the activity of the stations.

2 Dataset description

We analyze Lyon's shared bicycling system *Vélo'v* on the basis of the data provided by JCDecaux, promotor and operator of the program. The dataset contains all the bicycle

Pierre Borgnat and Antoine Scherrer
CNRS, Université de Lyon, ENS Lyon, Laboratoire de Physique (UMR 5672), France

Eric Fleury
Université de Lyon, ENS Lyon, INRIA/D-NET, LIP (UMR 5668 CNRS)

Céline Robardet (Corresponding Author)
Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France

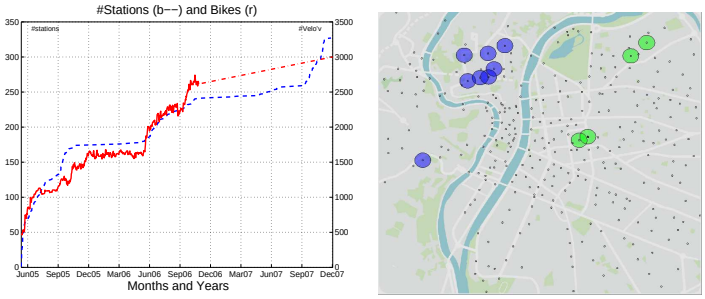


Fig. 1 Global time evolution of the system and spatial repartition. (a) Evolution of the number of existing stations (left side) and bikes (right side) during the studied time period (May 2005 - January 2008). (b) Geographical location of stations in Lyon and Villeurbanne, including anisotropic stations (as commented in Section 3).

trips that occurred between the 25th of May 2005 and the 12th of December 2007. Each record is anonymized and is made of the information about the date and time of the beginning of the trip, and of its end and the IDs of the departure and arrival stations (their geographical location being known).

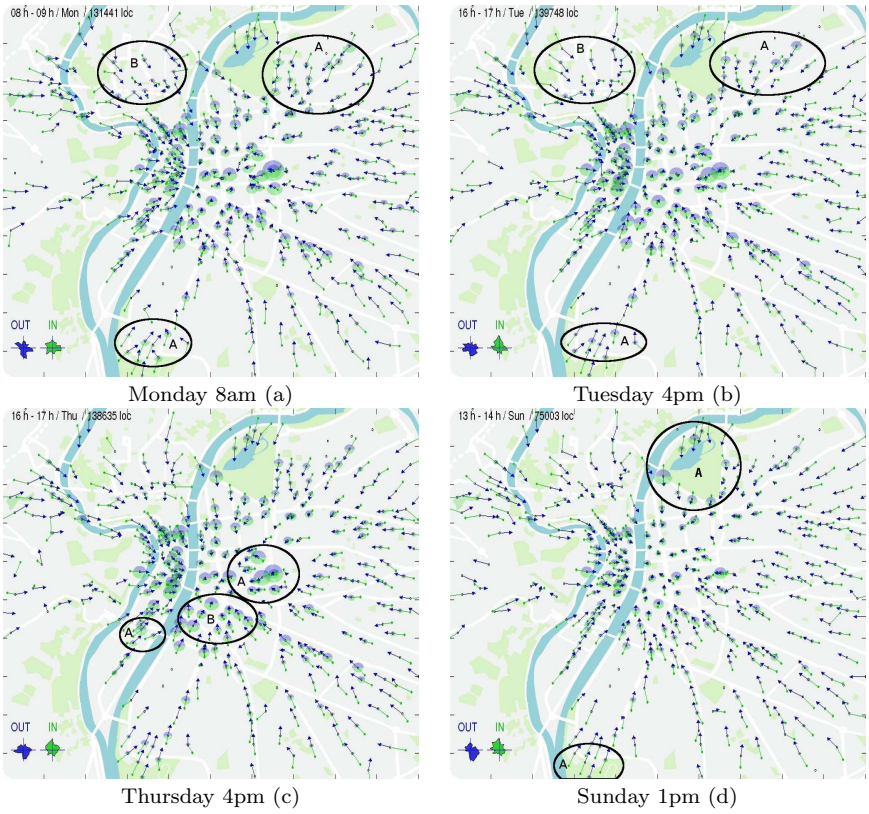
During this period, there were more than 13 millions hired bicycle trips. An important characteristic is that this bicycle program was in expansion. The *Vélo'v* system opens officially on May 19th 2005 and stations (resp. bikes) have been introduced continuously during the take off and lifetime of the system. Fig. 1(a) depicts the load of the systems (station being open/equipped regularly between May 2005 and October 2005). After this period, the curve reaches a plateau (October 2005 to May 2006) before a new phase of expansion that ends in January 2008 where the total current number of installed stations was reached (338 stations).

From this dataset, a global behavior of the program was analyzed by looking at the number of hourly hiring and by proposing a statistical model for it [4]. In the present work, the focus is on the spatial patterns of activity in the city, by studying how the rides are distributed onto the different stations. We first present a general analysis of the spatial patterns of activities of the stations. Then we turn to clustering of sets of pairwise flows between the stations that will reveal how each station's activity is correlated to varying social activities in time.

3 Spatial patterns of activities at the stations

Vélo'v network is not a self-regulated network since some stations have an incoming and leaving traffic unbalanced. This asymmetry is corrected by a small number of trucks equipped with trailers which move bicycles from one station to another to balance the distribution on the network. We identify the stations that suffer from an anisotropic traffic as those whose difference between their number of entering and leaving trips is larger than three times the standard deviation of the distribution of these values.

Among the 12 anisotropic stations (see Fig. 1(b) highlighted stations), 8 have more leaving trips (dark blue circles) and are located on the top of the two hills that surround Lyon, and the 4 remaining anisotropic stations (light green circles) have more entering trips and are located near the biggest shopping center, the central railway station and on the university campus (where most of the student accommodations are).



Monday 8am (a) Tuesday 4pm (b)

Thursday 4pm (c) Sunday 1pm (d)

Fig. 2 Visualization of the traffic at all station. The mean direction and the amount of incoming and leaving traffic is shown (as explained in the text, Section 3).

On Fig. 2, the main characteristics of the traffic of each station are visualized: the average of the directions of incoming trips is represented with a light green segment whose length is proportional to the anisotropy of the station. Blue arrows represent the same information for leaving trips. The two semi-circles on each station have an area proportional to the number of in/out bicycles. Finally, on the bottom of each figure the histograms of the in and out directions over the whole city are plotted.

We observe main tendencies among the bicycle users. Zone A on Fig. 2 (a) and (b) corresponds to campus, and on Monday 8 am, those stations host many bikes whereas on Tuesday 4pm, they are mostly in deficit. Zone B corresponds to stations that are on the top of a hill and they are in deficit all along the day. Zone A on Fig.2 (c) corresponds to railway stations. Many people seem to leave a Vélo'v near one of the train stations to take a train on Thursday 4pm. Zone B, Fig.2 (c), is a residential area and there is an entering flow in the area on Thursday around 4pm. Zone A in Fig.2 (d) matches with the two largest parks of Lyon. On Sundays, many people have recreation time there. These first simple diagrams based on temporal pattern of each stations allow to differentiate the type of place. Some stations (zone A on Fig. 2 (c)) act like hubs. Other stations are more "recreational" and does not reveal rush activity peak and are more much more balanced that stations that offers a "commuter" pattern.

4 Analysis of flow activity between stations

A new approach based on spatial clustering of the flows between stations is proposed here, highlighting the main features of bicycle movements all along the week. Our approach is composed of three main steps. First, we identify time stamps where the traffic is the most important. Second, once known these time stamps, we classify the pairs of stations based on the correlation on the number of trips between the two stations along these key time stamps. Finally, we propose a way to visualize the most important flows. The high-dimensional characteristic of Vélo'v traffic data are unfavor-

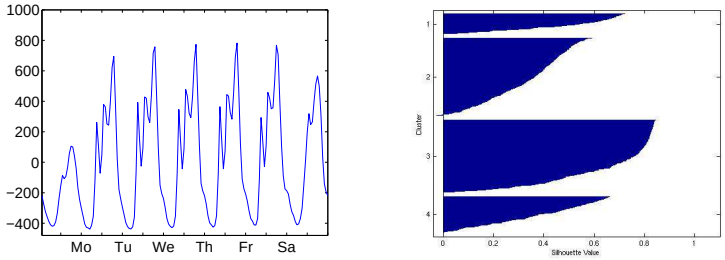


Fig. 3 Elements for the clustering of Flows. (a) Left: Projections of the times stamps on the first PCA axis. (b) Right: K-means silhouettes.

able to their analysis and visualization. To cope with this problem, we use Principal Component Analysis as a means to reduce the number of time stamps to consider. PCA transforms the original attributes, many of them being correlated, into a set of non-correlated components that are linear combinations of the original variables. The extracted non-correlated components are called Principal Components (PC) and have the important property of minimizing the squared error in reconstructing original data. By selecting components that account for most of the variation in the original multivariate data, PCA enables to summarize the data with little loss of information.

For the two years traffic, we aggregate the flows between stations per day of week and hour— calling this a time stamp. To isolate the most important ones, we compute PCA on the matrix (time stamps $\times$ couple of stations), keeping only the couples of stations whose maximum number of travels during a time stamp is greater than 45. Fig. 3(a) shows the projections of the time stamps on the first PCA axis which gathers 54.6% of the variance and look alike the aggregated number of hiring par hour (as found in [4]). It reflects the temporal patterns along the week showing where the number of rides is the larger: most Saturdays and Sundays rentals are held during 1pm-2pm and 5pm-6pm; for the 5 working days, we observe peak during 8am-9am, 12am-1pm and 6pm-7pm. The next two components gather respectively 13.4% and 5.6% of the variance and are neglected for this analysis. Hence, 19 most important time stamps are identified from this PCA component, corresponding to local maxima in the activity. We select those 19 times stamps to describe the activity between pairs of stations.

The second step of the analysis is to apply K-means algorithm on the selected data in order to synthesize the main trends of the rental activity on the network. The distances between every couples of pairs of stations were evaluated by the correlation between the temporal vectors of number of rentals.

Silhouette measures the quality of a clustering by estimating how similar a pair of stations is to pairs in its own cluster vs. pairs in other clusters, and ranges from

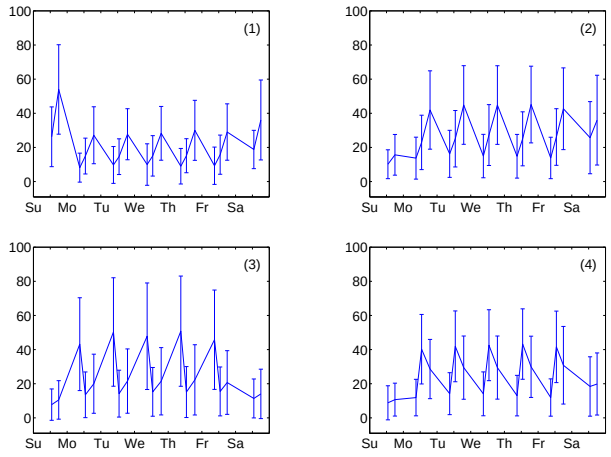


Fig. 4 Mean and standard deviation of the number of rentals for clusters (1) to (4).

-1 to 1. It is defined as $S(i) = \frac{\min_k(\overline{d_B}(i,k)) - \overline{d_W}(i)}{\max(\overline{d_W}(i), \min_k(\overline{d_B}(i,k)))}$, where $\overline{d_W}(i)$ is the average distance from the i-th point to the other points in its own cluster, and $\overline{d_B}(i,k)$ is the average distance from the i-th point to points in another cluster k. K-means finds 4 well separated clusters whose silhouette values of pairwise stations is shown on Fig. 3(b). Pairs of stations are closer to the ones of the same cluster than to pairs of others clusters, if one excepts 6 pairs among 1046 – this attests of the quality of the clustering.

Let us now identify to what the clusters correspond. Fig 4 shows the mean and standard deviation of the number of rentals for each cluster. The peaks of activity of each cluster are easily identifed: cluster 1 corresponds to rentals on Sundays at noon and more importantly at 5pm; the other clusters correspond to rentals made mostly on working days: cluster 2 corresponds to travels at 6pm, except on Fridays at 5pm; cluster 3 corresponds to hirings at 9am and cluster 4 gathers travels around noon.

Fig. 5 locates on the map of Lyon the pairs of stations that are part of each cluster. To make the picture understandable, we grouped nearby stations and plot a line between two areas if there exists at least one station in each area forming a pair that belong to the corresponding cluster. Blue discs are graphical representation of those areas. The trips in clusters 1 are mainly along the two rivers and around the main parks of the city (in the north and the south of the map). We can also observe some travels between the university campus and the center of the city (North-east and the land between the two rivers). Cluster 2 and 3 are quite similar: they correspond to commutations to (cluster 3) and from (cluster 2) work. We can identify the main hubs of the network (train stations, campus, historical center, etc.). Cluster 4 is less dense than the two previous ones and includes travels related to the lunch break rides. It is worth noting that these clustering results are very stable: very similar results are obtained when applying the same methodology on trips of each single month.

5 Conclusion

In this paper we have introduced the notion of spatial analysis of community shared bicycle program based on the digital footprint of the system. Our main objectives were

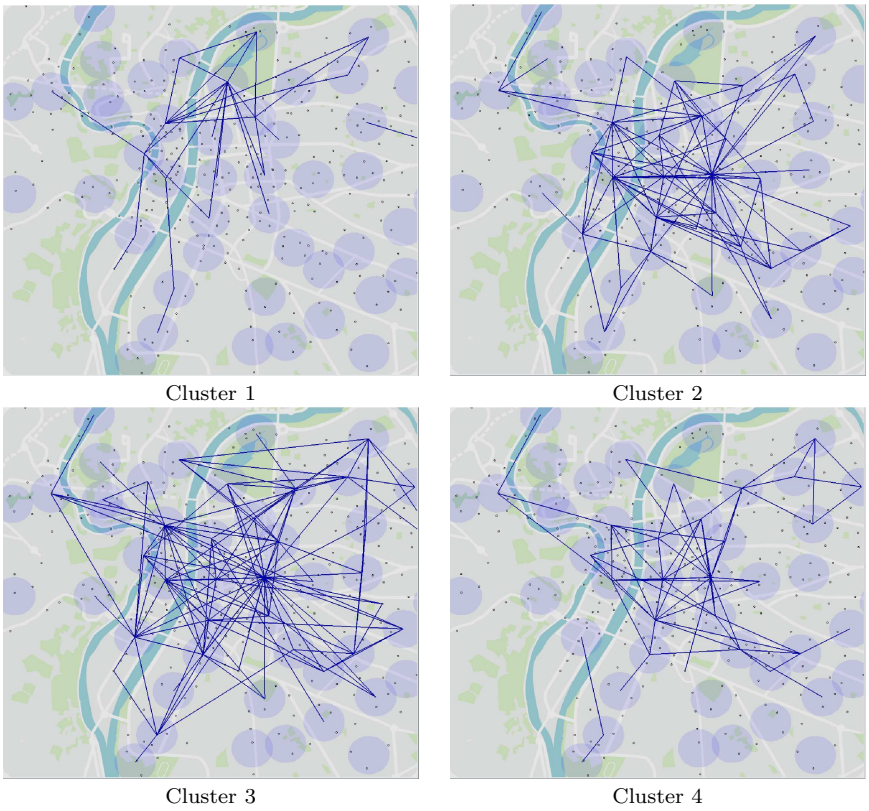


Fig. 5 Visualization of the 4 clusters of similar pairs of bicycle stations (by areas).

to use such new data to better understand the dynamics of the city and of human activity, and to introduce statistical analysis and tools suitable for spatio-temporal data produced by dynamic transportation networks. Open questions remain. We are in contact with JCDecaux and with the city Hall and they have complementary objectives: JCDecaux aims at optimizing the systems in terms of bike removal/balancing whereas from Lyon administration point of view, the main goal is to certify that the quality of service is achieved. Our preliminary analysis will be extended to address such issues.

Acknowledgments

This work was made by possible by the kind help of JCDecaux and the Grand Lyon. This work is part of an on-going project of scientific studies about the Vélo'v program and is supported by the IXXI (Institut des Systèmes Complexes – Complex Systems Institute) of Lyon. The authors would like to thank Jean-Baptiste Rouquier and Pablo Jensen for interesting discussions in this project.

References

1. http://www.velov.grandlyon.com/
2. Girardin, F. "Revealing Paris Through Velib' Data" http://liftlab.com/think/fabien/2008/02/27/revealing-paris-through-velib-data/ (2008).
3. Froehlich, J., Neumann, J., and Oliver, N. "Measuring the Pulse of the City through Shared Bicycle Programs" *Int. Workshop UrbanSense08*, Raleigh, North Carolina, USA (2008).
4. Borgnat, P., Abry, A., and Flandrin, F., "Studying Lyon's Velo'V: A Statistical Cyclic Model Vélo'v", Submitted Preprint (March, 2009).